UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

January 13, 2022

MEMORANDUM FOR Carolyn M. Pickering
Survey Director, Survey of Income and Program Participation

From: Anthony G. Tersine, Jr. *Anthony G. Tersine, Jr.*
Chief, Demographic Statistical Methods Division

Subject: Source and Accuracy Statement for The Survey of Income and Program Participation Calendar Year 2018 Data Collection Public Use Files (S&A-23)

This memorandum includes the Source and Accuracy Statement for the Survey of Income and Program Participation (SIPP) data collected during Calendar Year 2018.

The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY22-POP001-0040.

If you have any questions about this document, please contact Tracy Mattingly at tracy.l.mattingly@census.gov or 301-763-6445 or Mahdi Sundukchi at mahdi.s.sundukchi@census.gov or 301-763-4228 or Faith Nwaoha-Brown at faith.n.nwaoha.brown@census.gov or 301-763-4696.

cc:

Elizabeth Sinclair	(ADDP)	Jonathan Rothbaum	
Kevin Tolliver		Bethany DeSalvo	
David Waddington	(SEHSD)	Laryssa Mykyta	
Jason Fields		Kurt Bauman	
Hyon Shin		Sharon Stern	
Adam Smith		Rachel Shattuck	
Trudi Renwick		James Farber	(DSMD)
Stephanie Galvin		Tracy Mattingly	
Alfred Gottschalck		Mahdi Sundukchi	
Rose Kreider		Ashley Westra	
Mark Klee		Faith Nwaoha-Brown	
Brian McKenzie		Julia Yang	
Liana Fox		Brice Gnahore	
Lindsay Monte		Jacob Eliason	

Record of Changes

Version	Date	Description of Change	Author
1.0	01/21/2021	Original release	Faith N Nwaoha-Brown Brice Gnahore Mahdi Sundukchi
1.1	01/13/2022	Added new table of calendar year 2017 SIPP key estimates computed from CY18 data (Table 5) and renumbered subsequent tables.	Faith N Nwaoha-Brown

Demographic Statistical Methods Division
Sample Design and Estimation

Source and Accuracy Statement for Calendar Year 2018 Data Collection of the Survey of Income and Program Participation (SIPP)

Version 1.1
January 13, 2022

Faith Nwaoha-Brown
Brice Gnahore
Mahdi Sundukchi

James Farber, ADC
Tracy Mattingly, Lead Scientist

DRB Clearance Number – CBDRB-FY22-POP001-0040

Table of Contents

List of Tables

[This page is intentionally left blank]

1. Data Collection and Estimation

1.1 Source of Data

The data were collected in the 2018 Panel of the Survey of Income and Program Participation (SIPP) [1]. The population represented in the 2018 SIPP (the population universe) is the civilian noninstitutionalized population living in the United States. The institutionalized population which is composed primarily of the persons in correctional institutions and nursing homes (94 percent of the four million institutionalized people in the 2010 Census) is excluded from the SIPP universe.

The SIPP 2018 Panel sample is located in 686 Primary Sampling Units (PSUs), each consisting of a county or a group of contiguous counties. Of these 686 PSUs, 252 are self-representing (SR) and 434 are non-self-representing (NSR). SR PSUs have a probability of selection of one and NSR PSUs have a probability of selection less than one. Within PSUs, Housing Units[2] (HUs) were systematically selected from the Master Address File (MAF), which is the U.S. Census Bureau's official inventory of known housing units. The MAF was created using the decennial censuses, as well as the U.S. Postal Service's Delivery Sequence File (DSF). The Census Bureau continues to update the MAF using the DSF and various automated, clerical, and field operations.

Housing Units were classified into two strata[3], such that one stratum had a higher concentration of low income households than the other. We oversampled the low income stratum by 29 percent to increase the accuracy of estimates for low income households and program participation. Analysts are strongly encouraged to use the SIPP weights when computing estimates since households are not selected with equal probability.

Each household in the SIPP 2018 Panel sample was scheduled to be interviewed at yearly intervals over a period of roughly four years. The reference period for the interview questions is the preceding twelve-month calendar year. The most recent month, December, is designated reference month 12 and the earliest month, January, is reference month one. One cycle of interviewing covering the entire sample, using the same questionnaire, is called a wave and the SIPP 2018 Panel is comprised of four waves. Wave 1 interviews were conducted from February through July of 2018, collecting data on January through December 2017. Data for up to 12 reference months are available for persons on the file. Specific months available depend on a person's sample entry and/or exit date.

The SIPP 2018 Panel will overlap with new panels that begin each subsequent year. Figure 1 depicts the overlapping panel design for SIPP 2018 and subsequent panels. Beginning in 2019,

---

[1] This source and accuracy statement can also be accessed through the U.S. Census Bureau website at http://www.census.gov/programs-surveys/sipp/tech-documentation/source-accuracy-statements.html

[2] The SIPP selects housing units which may or may not be occupied at the time of interview; occupied housing units are referred to as households.

[3] Household income used to determine the low income and non-low income strata was obtained from the most recent American Community Survey (ACS) data that was available when the SIPP sample was selected in 2017.

sampled households from multiple panels will be interviewed concurrently during each annual cycle and the combined data will be referred to by the calendar year of collection. For example, SIPP 2018 Wave 2 and SIPP 2019 Wave 1 interviews occurred in 2019 and covered the same reference period – January to December 2018. The obtained data will be combined, edited, and published as *SIPP Calendar Year 2019 data*.

The overlapping panel design was first adopted during the early 1990s SIPP panels – SIPP 1990 through SIPP 1993 panels. It allows analysts to combine data from multiple panels covering the same reference period for cross-sectional analyses, thus increasing the sample size and decreasing standard errors of cross-sectional estimates (U.S. Census Bureau, 2020). Section 1.5 of the *2018 SIPP Users' Guide* further details the SIPP overlapping panel design, including comparisons to previous SIPP Panels. The remainder of this document focuses on the SIPP 2018 Panel.

Figure 1: Overlapping Panel Illustration for the 2018 and subsequent Panels.

<table>
<tr><td></td><td colspan="10">Year of interview</td></tr>
<tr><td rowspan="8">Panel</td><td>2017</td><td>2018</td><td>2019</td><td>2020</td><td>2021</td><td>2022</td><td>2023</td><td>2024</td><td>2025</td><td>…</td></tr>
<tr><td></td><td colspan="4">2018 Panel</td><td colspan="5"></td></tr>
<tr><td colspan="2"></td><td colspan="4">2019 Panel</td><td colspan="4"></td></tr>
<tr><td colspan="3"></td><td colspan="4">2020 Panel</td><td colspan="3"></td></tr>
<tr><td colspan="4"></td><td colspan="4">2021 Panel</td><td colspan="2"></td></tr>
<tr><td colspan="5"></td><td colspan="4">2022 Panel</td><td></td></tr>
<tr><td colspan="6"></td><td colspan="4">2023 Panel</td></tr>
<tr><td colspan="7"></td><td colspan="3">2024 Panel</td></tr>
</table>

In Wave 1, the SIPP 2018 Panel began with a sample of about 53,500 HUs. About 8,800 of these HUs were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. Field Representatives (FRs) were able to obtain interviews from about 26,000 of the eligible households. FRs were unable to interview approximately 18,500 eligible households in the panel because the occupants: (1) refused to be interviewed; (2) could not be found at home; (3) were temporarily absent; or (4) were otherwise unavailable. Thus, occupants of about 58 percent of all eligible households participated in the first interview of the 2018 Panel.

Only original sample people (interviewed persons in Wave 1 sample households) and people living with them are eligible to be interviewed in subsequent waves. Original sample people who move from their Wave 1 address to a new address in later waves are still included in the SIPP sample. However, FRs attempt telephone interviews in lieu of in-person interviews if their new address is more than 100 miles from a SIPP sample area.

Since the SIPP follows all original sample members, those that form new households are also included in the SIPP sample. This expansion of original households can be estimated within the interviewed sample but is impossible to determine within the noninterviewed sample.

Therefore, a growth factor based on the growth in the known sample is used to estimate the unknown expansion of the noninterviewed households. Growth factors account for the additional nonresponse stemming from the expansion of noninterviewed households and are calculated for Wave 2 and later waves. They provide a more accurate estimate of the weighted counts of noninterviewed households at each wave.

There are two categories of noninterviewed households: Type A and Type D. Type A noninterviewed households are eligible households where the interviewer obtains no interview. Type D noninterviewed households are previously interviewed households who moved to an unknown address or moved more than 100 miles from a SIPP interviewer and no telephone interview could be conducted. As a result, Type D noninterviews only occur from Wave 2 onwards. To calculate this loss of sample, or "sample loss," we use Formula (1):

$$Sample\ Loss = \frac{(A_1 \times GF_c) + A_c + D_c}{I_c + (A_1 \times GF_c) + A_c + D_c} \qquad (1)$$

where:

A_1 = weighted number of Type A noninterviewed households in Wave 1
A_c = weighted number of Type A noninterviewed households in the current wave
D_c = weighted number of Type D noninterviewed households in the current wave
I_c = weighted number of interviewed households in the current wave
GF_c = growth factor associated with the current wave.

The SIPP 2018 Wave 1 weighted sample loss was calculated using equation (1) and tabulated in Table 1. The distribution of Type A nonrespondents by nonresponse reason is shown in Table 2.

Table 1. SIPP 2018 Panel Household Counts, Sample Loss, and Weighted Response Rates

Wave	**Eligible Households[1]**	**Interviewed Households**	**Type A Households**		**Type D Households**		**Growth Factor**	**Cumulative Weighted Response Rates (percent)**	**Weighted Sample Loss (percent)**
			Total	**Weighted Rate (percent)**	**Total**	**Weighted Rate (percent)**			
1	45,000	26,000	18,500	41.56	–	–	–	58.44	41.56

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation
[1] Interviewed and noninterviewed households may not sum up to eligible households due to rounding.

Table 2. SIPP 2018 Panel Components of Type A Nonresponse by Wave

Wave	Language Problem	Unable to Locate	No One Home	Temporarily Absent	Household Refused	Other
1	1.450	0.210	13.29	2.24	76.14	6.650

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

1.2 Weights Produced

The SIPP produces three weights for cross-sectional and longitudinal analyses: monthly weights (*WPFINWGT*), calendar year weights (*CY<year>*), and panel weights (*FINPNL<wave>*). Monthly and calendar year weights are cross-sectional weights because they are computed using data obtained from a single interview. Panel weights, however, are computed by combining data from multiple interviews – i.e. two or more waves – and therefore are longitudinal weights[4].

Monthly weights are used to calculate estimates for each of the 12 months within a wave. For each reference month, respondents who were in the SIPP survey universe and for whom data were obtained, receive positive monthly weights.

Calendar year weights cover the reference period from January to December of a specified calendar year and can be used to calculate estimates for any period within the year. For example, calendar year weights for Wave 1 of the SIPP 2018 Panel can be used to compute monthly[5], quarterly, and annual estimates for any time between January and December 2017.

Calendar year weights for the SIPP 2018 Panel are based on the SIPP survey universe in December of a designated year. All interviewed persons in the SIPP survey universe who have positive December monthly weights are assigned calendar year weights equal to their monthly December weights, regardless of their monthly interview status[6] in preceding months. As a result, separate calendar year weight files are not produced for the SIPP 2018 Panel.

Panel weights cover reference periods consisting of two or more waves – from the beginning of Wave 1 to the end of the Wave 2, 3, or 4 – and can be used to calculate estimates in this interval. Therefore, longitudinal weights are produced at the completion of the second and later waves. The eligible sample cohort for SIPP 2018 panel weights consists of persons who have positive monthly weights in December 2017. The monthly interview statuses of these persons are tracked from December 2017 to the last month of the reference period. Eligible persons are then classified as *interviewed for panel weights* and assigned positive longitudinal

[4] Panel weights and longitudinal weights are used interchangeable throughout this document.

[5] We recommend analysts use monthly weights instead of calendar year weights to compute monthly estimates as using calendar year weights may result in fewer sample persons.

[6] The monthly interview status for each month of a reference period indicates whether data were obtained – either reported by the respondent during an FR's interview or imputed – for a respondent in the SIPP survey universe for that month.

weights, if data were obtained for all subsequent months of the reference period, except for months in which they were *survey universe leavers*.

SIPP survey universe leavers for a given month are defined as sample persons who are known to have died or moved to an ineligible address including: institutions, military barracks, and non-US addresses. Eligible persons for whom data were not obtained for one or more months between December 2017 and the last month of the reference period, and were not survey universe leavers in those months are categorized as *noninterviewed for panel weights* and receive zero panel weights.

Table 3 specifies the reference period for calendar and panel weights in the SIPP 2018 Panel. Note that the calendar year weights for years 2018 (*CY2018*), 2019(*CY2019*), and 2020(*CY2020*) incorporate data from the SIPP 2018 through SIPP 2021 Panels in accordance with the overlapping panel design described in Section 1.1 and Figure 1.

Table 3. Calendar Year and Panel Weights' Reference Periods for the 2018 Panel

Variable Name	**Control Month**	**Beginning Wave**	**Beginning Month**	**Ending Wave**	**Ending Month**
			Calendar year weights		
CY2017	December 2017	Wave 1	January 2017	Wave 1	December 2017
CY2018[1]	December 2018	Wave 2	January 2018	Wave 2	December 2018
CY2019[2]	December 2019	Wave 3	January 2019	Wave 3	December 2019
CY2020[3]	December 2020	Wave 4	January 2020	Wave 4	December 2020
			Panel weights		
FINPNL2	December 2017	Wave 1	January 2017	Wave 2	December 2018
FINPNL3	December 2017	Wave 1	January 2017	Wave 3	December 2019
FINPNL4	December 2017	Wave 1	January 2017	Wave 4	December 2020

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation.

[1] CY2018 calendar year weights incorporate data from SIPP 2018 Wave 2 and SIPP 2019 Wave 1 panels.

[2] CY2019 calendar year weights incorporate data from SIPP 2018 Wave 3, SIPP 2019 Wave 2, and SIPP 2020 Wave 1 panels.

[3] CY2020 calendar year weights incorporate data from SIPP 2018 Wave 4, SIPP 2019 Wave 3, SIPP 2020 Wave 2, and SIPP 2021 Wave 1 panels.

1.3 Estimation

The SIPP estimation procedure involves several stages of weight adjustments to derive the final person level weights. For Wave 1 cross-sectional weights, i.e. monthly and calendar year weights, each eligible household is first given a base weight (BW) equal to the inverse of its probability of selection. Next, a weighting control factor (WCF) is used to adjust for subsampling done in the field when the number of sampled units is much larger than expected. A noninterview adjustment factor (F_{N1}) is then applied to account for eligible households that FRs could not interview in Wave 1 to create the Wave 1 noninterview adjusted weights.

Household noninterview adjusted weights are assigned to each member of the household and a second stage adjustment factor (F_{2S}) is applied to determine monthly final person weights in Wave 1. The second stage adjustment equalizes married spouses' weights and rakes the sum of person weights to independent population controls (benchmark population estimates) by age, race, sex, Hispanic origin, and state of residence. The final cross-sectional weight (FW_c) for each month c, in Wave 1 is $FW_c = BW * WCF * FN_1 * F_{2S}$. Additional details of the weighting process are in Tersine (2020a).

Each eligible household in subsequent waves, i, is assigned an initial weight equal to its Wave 1 household noninterview adjusted weight. Initial weights are multiplied by movers' adjustment factors to account for multiple chances of selection of movers[7]. The resulting movers' weights are multiplied by the noninterview adjustment factor (F_{Ni}) to adjust for household nonresponse in the current wave and create household noninterview adjusted weights, which are assigned to all household members. Finally, the noninterview adjusted weights are raked to both national and state level controls for each of the twelve calendar months in the wave and married spouses' weights are equalized during second stage adjustments to determine final person monthly weights.

The weighting procedure for overlapping panel data (i.e. pooled data from two or more panels whose sample persons were interviewed in the same year as shown in Figure 1, and hence have the same reference period) is divided into two steps. First, the household noninterview adjustment is conducted separately for each panel due to the different response rates across panels. Next, the second stage adjustment is implemented on the pooled sample from all contributing panels. The second stage adjustment for overlapping panels incorporates an *overlapping panel factor* to adjust for combining sample households from multiple panels; equalizes married spouses' weights; and rakes persons weights to independent population controls for each of the 12 reference months to determine final person weights for all interviewed persons (Tersine, 2020b).

[7] Movers – persons who move into SIPP sample households after Wave 1 interviews – have two chances to become SIPP sample persons: (a) selection into original SIPP sample households in Wave 1 or (b) selection by moving into a sample household after Wave 1.

1.4 Population Controls

The 2018 SIPP estimation procedure adjusts weighted sample estimates to agree with independently derived population estimates of the civilian noninstitutionalized population. This attempts to correct for undercoverage and thereby reduces the mean square error of the estimate. The national and state level population controls are obtained directly from the Population Division and are prepared each month to agree with the most current set of population estimates released by the Census Bureau's population estimates and projections program.

The national level controls are distributed by demographic characteristics as follows:

- Age, Sex, and Race (White alone, Black alone, and all other groups combined)
- Age, Sex, and Hispanic Origin

The state level controls are distributed by demographic characteristics as follows:

- State, Age, and Sex
- State, Hispanic origin
- State, Race (Black alone, all other groups combined)

The estimates begin with the latest decennial census as the base and incorporate the latest available information on births and deaths along with the latest estimates of net international migration. The net international migration component in the population estimates includes a combination of:

- Legal migration to the U.S.,
- Emigration of foreign born and native people from the U.S.,
- Net movement between the U.S. and Puerto Rico,
- Estimates of temporary migration, and
- Estimates of net residual foreign-born population, which include unauthorized migration.

Because the latest available information on these components lags the survey date, it is necessary to make short-term projections of these components to develop the estimates for the survey date.

1.5 Use of Weights

The SIPP 2018 Panel monthly, calendar year, and panel weights are produced at the person level and intended for analyzing data at the person level. Every interviewed person in the SIPP universe for a given reference month – and for whom data were obtained – has a person month weight for that reference month. Likewise, persons who had data for the month of December have calendar year weights, and persons categorized as *interviewed for a longitudinal reference perio*d are assigned corresponding panel weights. Chapter 7 of the 2018 *SIPP User's Guide* provides additional information on how to use the survey weights.

In previous SIPP panels prior to the 2014 SIPP, public use files also contained household, family, and related subfamily monthly weights for analyzing the data at the appropriate household and family levels. These weights were set to be the person month weight of the household, family, or subfamily reference person[8] for that reference month[9]. For the SIPP 2014 Panel and subsequent panels, the household structure of an interviewed unit is only set for the interview month. Up to five addresses are recorded for each person for the reference period, so interviewed persons can live in different households depending on the reference month. Therefore, for each reference month it is possible to tell which interviewed persons lived together and their relationships to each other, but the files do not specify a household ID or reference person for each of the reference months. The same is true for families. If a data user would like to conduct analysis at the household or family level, person weights can be used to specify a single household or family weight. One option is to use the average of the person month weights for all persons in the household or family. Another option is to specify a household or family reference person and use his or her person month weight as the household or family weight.

All estimates may be divided into two broad categories: longitudinal and cross-sectional. Longitudinal estimates require that data records for each person be linked across interviews, whereas cross-sectional estimates do not. For example, estimating the average duration of unemployment in the 2018 fiscal year from October 2017 to September 2018 requires linking records from Wave 1 and Wave 2 of the 2018 SIPP Panel and would be a longitudinal estimate. Cross-sectional estimates can combine data from different interviews only at the aggregate level because there is no linkage between interviews; for example, comparing the unemployment rate among all persons interviewed in September 2017 to that of all persons interviewed in September 2018.

Data users are strongly encouraged to use the weights provided as instructed. Weights can vary due to the sample design, the level of nonresponse, and the time period of the estimate. Calculating estimates without using the weights as instructed will produce erroneous results.

Some basic types of estimates that can be constructed using monthly, calendar year, and panel weights are described below in terms of estimated numbers. More complex estimates, such as percentages, averages, ratios, etc., can be constructed from the estimated numbers.

1. The number of people who have ever experienced a characteristic during a given time period.

[8] During interviews, FRs identify a reference person who is usually the owner or renter of the residence. A household's reference person may change from wave to wave due to changes in the household composition overtime.
[9] Only person-level calendar year and panel weights were available in previous SIPP panels.

To construct such an estimate, use the person weight for the shortest time period which covers the entire time period of interest. Then sum the weights over all people who possessed the characteristic of interest at some point during the time period of interest. For example, to estimate the number of people who received Supplemental Nutrition Assistance Program (SNAP) benefits in January 2017, sum the monthly weights (*WPFINWGT*, with *monthcode*=1) of all persons who received SNAP benefits in January 2017.

2. The amount of a characteristic accumulated by people during a given time period.

   To construct such an estimate, use the person weight for the shortest time period which covers the entire time period of interest. Compute the product of the weight and the amount of the characteristic and sum this product over all appropriate people. For example, to estimate the aggregate 2017 annual income of people who were employed during all 12 months of the year, multiply the 2017 calendar year weights (*WPFINWGT*, with *monthcode*=12) of all persons who were employed in all months of 2017 by their annual income and sum the resulting products.

3. The average number of consecutive months of possession of a characteristic (i.e., the average spell length for a characteristic) during a given time period.

   For example, one could estimate the average length of each spell of Supplemental Security Income (SSI) receipt during 2017. One could also estimate the average spell of unemployment that elapsed before a person found a new job. To construct such an estimate, first identify the people who possessed the characteristic at some point during the time period of interest. Then create two sums of these persons' appropriate weights: (1) sum the product of the weight times the number of months the spell lasted and (2) sum the weights only. The estimated average spell length in months is computed as (1) divided by (2). A person who experienced two spells during the time period of interest would be treated as two people and appears twice in sums (1) and (2). An alternate method of calculating the average can be found in the section "Standard Error of a Mean."

4. The number of month-to-month changes in the status of a characteristic (i.e., number of transitions) summed over every set of two consecutive months during the time period of interest.

   To construct such an estimate, sum the appropriate person weights each time a change is reported between two consecutive months during the time period of interest. For example, to estimate the number of people who were in poverty in July 2017 and transitioned out of poverty in August 2017, add the Wave 1 calendar year weights of each person who had such a change. To estimate the number of changes in monthly earned income during the third quarter of 2017, sum together the estimate of the weighted number of people whose earned income varied between July and August, between August and September, and between September and October.

Note that spell and transition estimates should be used with caution because of the biases that are associated with them. Sample people tend to report the same status of a characteristic for all months of a reference period. This tendency also affects transition estimates in that, for many characteristics, the number of characteristics, the number of month-to-month transitions reported between the last month of one reference period and the first month of the next reference period are much greater than the number of reported transitions between any two months within a reference period. Additionally, spells extending before or after the time period of interest are cut off (censored) at the boundaries of the time period. If they are used in estimating average spell length, a downward bias will result.

5. Monthly estimates of a characteristic averaged over a number of consecutive months.

For example, one could estimate the monthly average number of Temporary Assistance for Needy Families (TANF) recipients from July 2017 through December 2017. To construct such an estimate, first form an estimate for each month in the time period of interest. Sum the Wave 1 calendar year weight, *CY2017* (i.e. *WPFINWGT* with *monthcode*=12), of all persons who possessed the characteristic of interest during each of the six months of interest. Then sum the monthly estimates and divide by six, the number of months.

2. Accuracy of Estimates

SIPP estimates are based on a sample; they may differ somewhat from the figures that would have been obtained from a complete census using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and nonsampling errors. For a given estimator, the difference between an estimate based on a sample and the estimate that would result if the sample were to include the entire population is known as sampling error. For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error.

2.1 Nonsampling Error

Nonsampling errors can be attributed to many sources:

- inability to obtain information about all cases in the sample
- definitional difficulties
- differences in the interpretation of questions
- inability or unwillingness on the part of the respondents to provide correct information

- errors made in the following: collection such as in recording or coding the data, processing the data, estimating values for missing data
- biases resulting from the differing recall periods caused by the interviewing pattern used and undercoverage.

Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP can be found in the *SIPP Quality Profile, 1998 SIPP Working Paper Number 230*, issued June 1998 (Kalton, 1998) and in the *SIPP 2018 Users' Guide* (U.S. Census Bureau, 2020).

Undercoverage in SIPP results from missed HUs and missed persons within sample HUs. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-sex-Hispanic origin population controls during the second stage adjustment step of the SIPP weighting procedure partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex-Hispanic origin group.

A common measure of survey coverage is the coverage ratio, the estimated population before second stage adjustment divided by the independent population control. Table 4 shows SIPP 2018 Panel coverage ratios for age-sex-race-Hispanic origin groups in December 2017 using calendar year cross-sectional weights prior to the second stage ratio adjustment. The SIPP coverage ratios exhibit some variability from month to month. Other Census Bureau household surveys (e.g. the Current Population Survey) experience similar coverage.

Caution should be exercised when comparing the SIPP 2018 Panel data with data from other SIPP products[10] or with data from other surveys. The comparability problems are caused by sources such as the seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the *SIPP Quality Profile* (Kalton, 1998) for known differences with data from other sources and further discussions.

[10] Analysts should be cautious when comparing estimates from the 2014 and subsequent SIPP panels to previous panels as the SIPP was redesigned following the SIPP 2008 Panel. The questionnaire, interview frequency, collection format, and sample size of the SIPP 2014 and later panels differ from those of previous SIPP panels. Detailed information on the reengineered SIPP are provided in the 2014 SIPP Users' Guide and the 2018 SIPP Users' Guide.

Table 4. Coverage Ratios for December 2017 for *CY2017* Weights by Age, Race, Sex, and Hispanic Origin

Age	All Persons		White non-Hispanic		Black non-Hispanic		Other non-Hispanic		Hispanic (of any race)	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
<15	0.86	0.85	0.91	0.86	0.80	0.85	0.79	0.88	0.82	0.81
15	0.81	0.91	0.81	0.94	0.87	0.93	0.81	0.83	0.76	0.85
16-17	0.86	0.90	0.89	0.93	0.87	0.91	0.82	0.84	0.82	0.86
18-19	0.88	0.86	0.92	0.87	0.83	0.91	0.81	0.84	0.87	0.82
20-21	0.88	0.84	0.93	0.87	0.75	0.77	0.82	0.85	0.86	0.79
22-24	0.79	0.75	0.81	0.75	0.73	0.78	0.80	0.84	0.77	0.69
25-29	0.76	0.82	0.80	0.83	0.66	0.69	0.87	0.84	0.66	0.88
30-34	0.82	0.83	0.88	0.85	0.69	0.68	0.85	0.83	0.73	0.86
35-39	0.83	0.88	0.88	0.89	0.69	0.88	0.87	0.81	0.75	0.90
40-44	0.84	0.85	0.86	0.84	0.83	0.87	0.87	0.80	0.76	0.87
45-49	0.87	0.88	0.89	0.89	0.89	0.80	0.86	0.95	0.77	0.85
50-54	0.90	0.93	0.92	0.94	0.98	0.87	0.85	1.01	0.81	0.90
55-59	0.93	0.97	0.95	0.98	0.87	1.00	1.03	0.90	0.83	0.93
60-61	0.98	0.98	1.01	0.99	0.86	1.00	1.03	0.92	0.85	0.95
62-64	1.00	1.02	1.04	1.04	0.85	1.00	1.01	0.91	0.89	0.99
65-69	1.03	1.06	1.05	1.08	0.97	1.06	1.09	1.00	0.88	0.92
70-74	1.07	1.10	1.09	1.14	0.95	1.07	1.11	0.99	0.92	0.95
75-79	1.11	1.11	1.14	1.14	0.95	1.05	1.12	0.98	0.91	0.97
80-84	1.18	1.01	1.21	1.02	0.99	1.06	1.13	0.99	1.02	0.85
85+	1.03	0.95	1.04	0.95	0.98	1.10	1.04	0.95	0.89	0.83

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation (SIPP)

3. Sampling Variability and Computation of Standard Errors

Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

3.1 Confidence Intervals

The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range about a given estimate that has a known probability of including the result of a complete enumeration. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate of all possible samples would include the actual value of the population parameter.

2. Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate of all possible samples would include the actual value of the population parameter.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate of all possible samples would include the actual value of the population parameter.

The estimate derived from all possible samples may or may not be contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the estimate derived from all possible samples is included in the confidence interval.

3.2 Hypothesis Testing

Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are 1) the population characteristics are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical. The Census Bureau's uses 90 percent confidence level for hypothesis testing but stricter confidence levels – for example 95 percent or higher – may also be used.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be S_{DIFF}.

Calculate a test statistic $|z|$, as $|z| = |X_A - X_B|/S_{DIFF}$. For large sample sizes, one can conclude the difference is significant at the 90 percent confidence level if $|z|$ is larger than the critical value of 1.645.

Confidence intervals can also be used to test for significant difference between two sample estimates. If $X_A - X_B$ is between $(-1.645 \times S_{DIFF})$ and $(+1.645 \times S_{DIFF})$, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand $X_A - X_B$, is smaller than $(-1.645 \times S_{DIFF})$ or larger than $(+1.645 \times S_{DIFF})$, the observed difference is significant at the 10 percent significance level. In this event, it is commonly accepted practice to say that the characteristics are significantly different. In accordance with the Census Bureau's Statistical Quality Standards, we recommend that users report only those differences that are significant at the 10 percent significance level or better. Of course, sometimes this conclusion will be wrong. When the characteristics are the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, at the 10 percent significance level, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously. A Bonferroni correction can be used to account for this potential problem. The Bonferroni method involves dividing your stated level of significance by the number of tests you are performing (Sedgwick, 2014; Stoline, 1981). This correction results in a conservative test of significance.

Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a small number of sample cases. We recommend a minimum weighted estimate – i.e. the weighted totals of the sample cases included in the analysis – of 150,800 for computing SIPP 2018 Panel Wave 1 estimates. Nonsampling error in one or more of the small number of cases providing the estimation can also cause large relative error in that particular estimate. Care must be taken in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

3.3 Calculating Standard Errors for SIPP Estimates

There are three main ways we calculate the Standard Errors (SEs) for SIPP Estimates. They are as follows:

- Direct estimates using replicate weight methods (highly recommended);
- Generalized variance function parameters (denoted as a and b); and
- Simplified tables of SEs based on the a and b parameters.

Replicate weight methods provide the most accurate variance estimates but may require more computing resources and expertise on the part of the user. The Generalized Variance Function (GVF) parameters provide a method of balancing accuracy with resource usage as well as having a smoothing effect on SE estimates across time. SIPP uses the Replicate Weighting Method to produce GVF parameters (see Chapter 7 of K. Wolter, *Introduction to Variance Estimation,* for more information). The GVF parameters are used to create the simplified tables of SEs provided in Tables 7 to 10. Table 5 shows calendar year 2017 SIPP key estimates, their standard errors, coefficient of variation, and item response rates computed from CY18 data. Standard errors for the key estimates were computed using replicate weights and Fay's BRR method.

Table 5. Calendar Year 2017 SIPP Key Estimates Computed from CY18 Data.

Key Estimate	Estimate	Standard Error	Coefficient of Variation	Item Response Rate
Median annual household earnings	42,120	491	1.17	66.78
Median annual household income	58,660	444	0.76	48.94
Median household net worth	103,900	1,734	1.67	32.02
Poverty rate	13.74	0.23	1.65	57.18
Percent of households receiving means-tested benefits[1]	19.21	0.40	2.06	79.97
Percent receiving means-tested transfer income[2]	13.73	0.20	1.47	86.11
Percent receiving Supplemental Nutrition Assistance Program (SNAP) benefits	11.04	0.20	1.80	94.16
Percent receiving Supplemental Security Income (SSI)	2.56	0.07	2.79	87.26
Percent receiving income from social insurance programs[3]	23.79	0.13	0.55	86.55
Percent receiving Old Age, Survivors and Disability Insurance (OASDI) income	21.17	0.12	0.55	87.46
Percent covered by public health insurance	36.19	0.26	0.73	86.08
Percent covered by Medicaid	20.62	0.26	1.25	87.66
Percent covered by Medicare	17.70	0.09	0.51	87.93
Percent covered by private health insurance	65.65	0.31	0.47	91.06
Percent covered by employer-sponsored health insurance	54.71	0.31	0.56	85.90

Key Estimate	Estimate	Standard Error	Coefficient of Variation	Item Response Rate
Percent uninsured	9.23	0.16	1.73	86.01
Percent with a retirement plan or account	42.49	0.24	0.56	86.25
Percent of workers working two or more jobs	15.06	0.22	1.45	88.15
Percent of workers who are self-employed	13.25	0.22	1.67	88.01
Percent of workers who worked full-time, year-round	62.16	0.30	0.49	81.96
Percent of workers who work from home at least one day per week	14.68	0.21	1.43	99.99
Children ever born per 1,000 women age 15 to 50	1,247	10.26	0.82	93.11
Children ever born per 1,000 men age 15 to 50	965	10.51	1.09	93.94
Percent of mothers with multi-partner fertility	18.83	0.28	1.46	69.30
Percent of fathers with multi-partner fertility	15.87	0.31	1.97	64.96
Percent with a disability (age 1+)	19.51	0.18	0.90	86.89
Percent of households who experienced a change in composition	15.42	0.23	1.50	100.00

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

[1] Means-tested benefits include Medicaid, Supplemental Nutrition Assistance Program (SNAP), Supplemental Security Income (SSI), General Assistance (GA), Temporary Assistance for Needy Families (TANF), or special Supplemental Nutrition Program for Women Infants and Children (WIC).

[2] Means-tested transfer income include SNAP, SSI, TANF, WIC, and GA.

[3] Social insurance programs include Old Age, Survivors and Disability Insurance (OASDI), Veteran Affairs (VA) benefits, Workers Compensation (WC), and Unemployment Compensation (UC).

3.4 Standard Error Parameters and Tables and Their Use

Most SIPP estimates have greater standard errors than those obtained through a simple random sample because of its two-stage clustered sample design. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required.

Estimates with similar standard error behavior were grouped together and two parameters (denoted as ***a*** and ***b***) were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These ***a*** and ***b*** parameters vary by characteristic and by demographic subgroup to which the estimate applies. Table 6 provides ***a*** and ***b*** parameters for the core domains to be used for the 2018 Panel cross-sectional (monthly and calendar year) estimates.

The creation of appropriate ***a*** and ***b*** parameters for the previously discussed types estimates are described below.

1. The number of people who have ever experienced a characteristic during a given time period.

   The appropriate ***a*** and ***b*** parameters are obtained directly from Table 6. The choice of parameter depends on the weights used, the characteristic of interest, and the demographic subgroup of interest.

2. Amount of a characteristic accumulated by people during a given time period.
   The appropriate ***b*** parameters are also taken directly from Table 6.

3. The average number of consecutive months of possession of a characteristic per spell (i.e., the average spell length for a characteristic) during a given time period.

   Start with the appropriate base ***a*** and ***b*** parameters from Table 6. The parameters are then inflated by an additional factor, *g*, to account for people who experience multiple spells during the time period of interest. This factor is computed by:

$$g = \frac{\sum_{i=1}^{n} m_i^2}{\sum_{i=1}^{n} m_i} \tag{2}$$

   where there are *n* people with at least one spell and m_i is the number of spells experienced by person i during the time period of interest.

4. The number of month-to-month changes in the status of a characteristic (i.e., number of transitions) summed over every set of two consecutive months during the time period of interest.

   Obtain a set of adjusted ***a*** and ***b*** parameters exactly as just described in 3, then multiply these parameters by an additional factor. Use 1.0 if the time period of interest is two months and 2.0 for a longer time period. (The factor of 2.0 is based on the conservative assumption that each spell produces two transitions within the time period of interest.)

5. Monthly estimates of a characteristic averaged over a number of consecutive months.

   Appropriate base ***a*** and ***b*** parameters are taken from Table 6. If one or more cross-sectional or longitudinal weight has been used in the monthly average (i.e., when Wave 2+ files are available), then there will be multiple GVF tables and a choice of parameters. Choose the table that gives the largest parameter.

For users who wish further simplification, we have also provided base standard errors for estimates of totals and percentages in Tables 7 through 10. Note that these base standard errors must be adjusted by a f factor provided in Table 6. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors for different estimates are given in the following sections. Later, we will describe how to use software packages to directly compute standard errors using replicate weights.

3.5 Standard Errors of Estimated Numbers

The approximate standard error, s_x, of an estimated number of persons, households, families, unrelated individuals and so forth, can be obtained in two ways. Note that neither method should be applied to dollar values.

The standard error may be obtained by the use of Formula (3):

$$s_x = f \times s, \tag{3}$$

where f is the appropriate f factor from Table 6, and s is the base standard error on the estimate obtained from Table 7 or 8.

Alternatively, s_x may be approximated by Formula (4):

$$s_x = \sqrt{ax^2 + bx} \tag{4}$$

Here x is the size of the estimate and a and b are the appropriate parameters from Table 6 associated with the characteristic being estimated (and the wave which applies). This formula was used to calculate the base standard errors in Tables 7 and 8. Use of Formula (4) will generally provide more accurate results than the use of Formula (3).

<u>Illustration 1.</u>
Suppose SIPP estimates based on Wave 1 of the 2018 panel show that there were 5,330,000 females aged 25 to 44 with a monthly earned income greater than $6,000 in September 2017. The appropriate parameters and factor from Table 6 are:

$$a = -0.00004128 \quad b = 5{,}546 \;\; f = 0.9980$$

Table 8 does not contain an entry for a sample size of 5,330,000; therefore, its corresponding base standard error s, in Formula (3) will be computed via linear interpolation.

First determine the sample size, x_1, (and its base standard error, y_1) on Table 8 that is closest to and also less than 5,330,000. Next determine the sample size, x_2, (and its base standard error,

y_2) on Table 8 that is closest to and also greater than 5,330,000. The linear interpolated base standard error s for a sample size x is defined as:

$$s = y_1 + \frac{(y_2 - y_1) \times (x - x_1)}{(x_2 - x_1)}$$

In our example, $x_1 = 5{,}000{,}000$, $y_1 = 165{,}550$, $x_2 = 7{,}500{,}000$, $y_2 = 201{,}953$ and $x = 5{,}330{,}000$

$$s = 165{,}550 + \frac{(201{,}953 - 165{,}550) \times (5{,}330{,}000 - 5{,}000{,}000)}{(7{,}500{,}000 - 5{,}000{,}000)} = 170{,}355$$

Using Formula (3), the approximate standard error is:

$$s_x = 0.998 \times 170{,}355 = 170{,}014\ females$$

Using Formula (4), the approximate standard error is:

$$s_x = \sqrt{(-0.00004128 \times 5{,}330{,}000^2) + (5{,}546 \times 5{,}330{,}000)} = 168{,}486\ females.$$

Using the standard error based on Formula (4), the approximate 90 percent confidence interval as shown by the data is from 5,052,841 to 5,607,159 females ($i.e.$, $5{,}330{,}000 \pm 1.645 \times 168{,}486$). Therefore, 90 percent of confidence intervals from all possible samples will contain the actual number of females aged 25 to 44 with a monthly earned income greater than \$6,000 in September 2017 in the population.

3.6 Standard Error of a Mean

A mean is defined here to be the average quantity of some item (other than persons, families, or households) per person, family or household. For example, it could be the average monthly household income for a specified demographic. The standard error of a mean can be approximated by Formula (5) below. Because of the approximations used in developing Formula (5), an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\bar{x}$ is:

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right) s^2}, \tag{5}$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

The population variance s^2 may be estimated by one of two methods. In both methods, we assume x_i is the value of the item for i^{th} unit. (A unit may be person, family, or household). To use the first method, the range of values for the item is divided into c intervals. The lower and upper boundaries of interval j are Z_{j-1} and Z_j, respectively. Each unit, x_i, is placed into one of c intervals such that $Z_{j-1} < x_i \leq Z_j$. The estimated population mean, $\bar{x}$, and variance, s^2, are given by the formulas:

$$\bar{x} = \sum_{j=1}^{c} p_j m_j$$

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \bar{x}^2 \tag{6}$$

where $m_j = (Z_{j-1} + Z_j)/2$, and p_j is the estimated proportion of units in the interval j. The most representative value of the item in the interval j is assumed to be m_j. If the interval c is open-ended, or no upper interval boundary exists, then an approximate value for m_c is

$$m_c = \frac{3}{2} Z_{c-1}.$$

In the second method, the estimated population mean, $\bar{x}$, and variance, s^2 are given by:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2 \tag{7}$$

where there are n units with the item of interest and w_i is the final weight for i^{th} unit. (Note that $\sum w_i = y$.)

Illustration 2.

Method 1

Suppose that based on Wave 1 data, the distribution of annual income for persons aged 25 to 34 who were employed for all 12 months of 2017 is given in Table A.

Table A. Hypothetical Distribution of Annual Cash Income among People 25 to 34 Years Old (Not Actual Data, Only Use for Calculation Illustrations)

Interval of Annual Cash Income	Number of People in Each Interval (in thousands)	Cumulative Number of People with at Least as Much as Lower Bound of Each Interval (in thousands)	Percent of People with at Least as Much as Lower Bound of Each Interval
under $5,000	370	23,527	100
$5,000 to $7,500	302	23,158	98.4
$7,500 to $9,999	447	22,856	97.1
$10,000 to $12,499	685	22,409	95.2
$12,500 to $14,999	935	21,724	92.3
$15,000 to $17,499	1,113	20,789	88.4
$17,500 to $19,999	1,298	19,675	83.6
$20,000 to $29,999	5,496	18,377	78.1
$30,000 to $39,999	4,596	12,881	54.7
$40,000 to $49,999	3,121	8,285	35.2
$50,000 to $59,999	1,902	5,164	21.9
$60,000 to $69,999	1,124	3,262	13.9
$70,000 and above	2,138	2,138	9.1

Using these data, the mean monthly cash income for persons aged 25 to 34 is $38,703. Applying Formula (6), the approximate population variance, s^2, is:

$$s^2 = \left(\frac{370}{23{,}527}\right)(2{,}500)^2 + \cdots + \left(\frac{2{,}138}{23{,}527}\right)(105{,}000)^2 - (38{,}703)^2 = 649{,}442{,}787$$

Using Formula (5) and a b parameter of 5,546 from Table 6, the estimated standard error of a mean $\bar{x}$ is:

$$s_{\bar{x}} = \sqrt{\frac{5{,}546}{23{,}527{,}000} \times 649{,}442{,}787} = \$391$$

Thus, the approximate 90 percent confidence interval as shown by the data ranges from $38,060 to $39,346.

<u>Method 2</u>

Suppose that we are interested in estimating the average length of spells of SNAP recipiency during the last six months of 2017, i.e. from July 2017 to December 2017, among the Black subpopulation. Also, suppose there are only 10 sample people in the subpopulation who were recipients of SNAP benefits. This example is a hypothetical situation used for illustrative purposes only; actually, 10 sample cases would be too few for a reliable estimate and the sum of weights (68,747) is lower than the recommended minimum value of 150,800. The number of consecutive months of SNAP recipiency during 2017 and 2017 calendar year weights (*CY2017*) are given in Table B below for each sample person:

Table B. Hypothetical Supplemental Nutrition Assistance Program (SNAP) Recipiency Spells from July 2017 to December 2017 among SIPP Sample Persons (Not Actual Data, Only Use for Calculation Illustrations).

Sample Person	Spell Length in Months	2017 Calendar Year Weight (*CY2017*)
1	2, 2	7,250
2	3	3,944
3	4	5,203
4	3, 2	5,786
5	6	3,549
6	6	5,335
7	4, 1	3,046
8	5	3,793
9	6	4,149
10	4	10,610

Using formula (7), the average spell of SNAP recipiency is estimated to be:

$$\bar{x} = \frac{(7{,}250)(2) + (7{,}250)(2) + \cdots + (10{,}610)(4)}{7{,}250 + 7{,}250 + \cdots + 10{,}610} = \frac{245{,}407}{68{,}747} = 3.570$$

The standard error will be computed by Formula (5). First, the estimated population variance can be obtained by Formula (7):

$$s^2 = \frac{(7{,}250)(2)^2 + (7{,}250)(2)^2 + \cdots + (10{,}610)(4)^2}{7{,}250 + 7{,}250 + \cdots + 10{,}610} - 3.570^2 = 2.347(months)^2$$

Next, the base b parameter of 5,696 is taken from Table 6 and multiplied by the factor computed from Formula (2):

$$g = \frac{2^2 + 1 + 1 + 2^2 + 1 + 1 + 2^2 + 1 + 1 + 1}{2 + 1 + 1 + 2 + 1 + 1 + 2 + 1 + 1 + 1} = 1.462$$

Therefore, the final b parameter is $1.462 \times 5{,}696 = 8{,}328$ and the standard error of the mean from Formula (5) is:

$$s_{\bar{x}} = \sqrt{\frac{(8{,}328)(2.347)}{68{,}747}} = 0.5332\, months$$

3.7 Standard Error of an Aggregate

An aggregate is defined to be the total quantity of an item summed over all the units in a group. The standard error of an aggregate can be approximated using Formula (8). As with the estimate of the standard error of a mean, the estimate of the standard error of an aggregate will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using Formula (6) or Formula (7) and b be the parameter associated with the particular type of item. The standard error of an aggregate is:

$$s_x = \sqrt{b \times y \times s^2}. \tag{8}$$

3.8 Standard Errors of Estimated Percentages

The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. For example, the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

There are two types of percentages commonly estimated. The first is the percentage of people sharing a particular characteristic such as the percent of people owning their own home. The second type is the percentage of money or some similar concept held by a particular group of people or held in a particular form. Examples are the percent of total wealth held by people with high income and the percent of total income received by people on welfare.

For the percentage of people, the approximate standard error, $s_{(x,p)}$, of the estimated percentage p can be obtained by the formula:

$$s_{(x,p)} = f \times s, \tag{9}$$

where f is the appropriate f factor from Table 6 and s is the base standard error of the estimate from Tables 9 and 10.

Alternatively, it may be approximated by the formula:

$$s_{(x,p)} = \sqrt{\frac{b}{x}(p)(100 - p)}, \tag{10}$$

from which the standard errors in Tables 9 and 10 were calculated. Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage $(0 < p < 100)$, and b is the parameter associated with the characteristic in the numerator. Use of Formula (10) will give more accurate results than use of Formula (9).

Illustration 3.

Suppose that using the 2017 calendar year weight, *CY2017*, it was estimated that 22,610,000 males were in poverty in January 2017 and an estimated 4.653 percent of them exited poverty in February 2017. Using Formula (10), with a b parameter of $5{,}696$ from Table 6, the approximate standard error is:

$$s_{(x,p)} = \sqrt{\frac{5{,}696}{22{,}610{,}000} \times 4.653 \times (100 - 4.653)} = 0.3343\ percent$$

Consequently, the 90 percent confidence interval as shown by these data is from 4.103 percent to 5.203 percent.

For percentages of money, a more complicated formula is required. A percentage of money will usually be estimated in one of two ways. It may be the ratio of two aggregates:

$$p_R = 100\left(\frac{x_A}{x_B}\right),$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_R = 100\left(\hat{p}_A\left(\frac{\bar{x}_A}{\bar{x}_B}\right)\right),$$

where x_A and x_B are aggregate money figures, $\bar{x}_A$ and $\bar{x}_B$ are mean money figures, and $\hat{p}_A$ is the estimated number in group A divided by the estimated number in group B. In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A\bar{x}_A}{\bar{x}_B}\right)^2\left[\left(\frac{s_p}{\hat{p}_A}\right)^2 + \left(\frac{s_A}{\bar{x}_A}\right)^2 + \left(\frac{s_B}{\bar{x}_B}\right)^2\right]}, \tag{11}$$

where s_p is the standard error of $\hat{p}_A$, s_A is the standard error of $\bar{x}_A$ and s_B is the standard error of $\bar{x}_B$. To calculate s_p, use Formula (10). The standard errors of $\bar{x}_B$ and $\bar{x}_A$ may be calculated using Formula (5).

It should be noted that there is frequently some correlation between $\hat{p}_A$, $\bar{x}_B$, and $\bar{x}_A$. Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

Illustration 4.
In 2017, 7.575 percent of households owned rental property. The mean value of rental property among these households who own them is $455,200, the mean value of assets among all households is $652,000 and the corresponding standard errors are 0.1627 percent, $23,970, and $19,280 respectively. Then the percent of all household assets held in rental property is:

$$100\left(0.07575 \times \frac{455{,}200}{652{,}000}\right) = 5.289\, percent$$

Using formula (11), the corresponding standard error is

$$s_i = \sqrt{\left(\frac{0.07575 \times 455{,}200}{652{,}000}\right)^2 \left[\left(\frac{0.001627}{0.07575}\right)^2 + \left(\frac{23{,}970}{455{,}200}\right)^2 + \left(\frac{19{,}280}{652{,}000}\right)^2\right]} = 0.339\, percent$$

3.9 Standard Error of a Difference

The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{{s_x}^2 + {s_y}^2 - r s_x s_y} \qquad (12)$$

where s_x and s_y are the standard errors of the estimates *x* and *y*.

The estimates can be numbers, percent, ratios, etc. The correlation between x and y is represented by *r*. The above formula assumes that the correlation coefficient between the characteristics estimated by *x* and *y* is non-zero. If no correlations have been provided for a given set of x and y estimates, assume *r* = 0. However, if the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration 5.
Supposed that SIPP estimates show 752,800 Hispanic males and 890,100 Hispanic females received Federal and/or State administered Supplemental Security Income (SSI) in December 2017. Then, using the *Male Hispanic* GVF parameters (a=-0.00019140 and b=5,598) and the Female Hispanic GVF Parameters (a=-0.00019190 and b=5,598) from Table 6 and Formula (4), the standard errors of these numbers are approximately 64,076 and 69,504 respectively. The difference in sample estimates is 137,300 and using Formula (12), the approximate standard error of the difference is:

$$\sqrt{64{,}076^2 + 69{,}504^2} = 94{,}533$$

Suppose that it is desired to test at the 10 percent significance level whether the number of female Hispanic SSI recipients was different from the number of male Hispanic SSI recipients in

December 2017. To perform the test, compute the Z statistic and compare it to the critical value of 1.645 for the 10 percent significance level.

$$Z = \frac{137{,}300}{94{,}533} = 1.452$$

Since 1.452 is less than 1.645, we can conclude that the number of male and female Hispanic SSI recipients are not significantly different at the 10 percent significance level.

We can also compare the difference of 137,300 to the product $1.645 \times 94{,}533 = 155{,}507$. Since the difference is less than 1.645 times the standard error of the difference, we come to the same conclusion that the number of SSI recipients in both groups are not significantly different at the 10 percent significance level.

3.10 Standard Error of a Median

The median quantity of some item such as income for a given group of people is that quantity such that at most 50 percent of the group have less and at most 50 percent of the group have more. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. To calculate standard errors on medians, the procedure described below may be used.

The median, like the mean, can be estimated using either data which have been grouped into intervals or ungrouped data. If grouped data are used, the median is estimated using Formulas (13) or (14) with $q = 0.5$, where q is the proportion of a group possessing characteristics of interest. If ungrouped data are used, the data records are ordered based on the value of the characteristic, then the estimated median is the value of the characteristic such that the weighted estimate of half – i.e. 50 percent – of the subpopulation falls at or below that value and half is at or above that value. Note that the method of standard error computation which is presented here requires the use of grouped data. Therefore, it should be easier to compute the median by grouping the data and using Formula (13) or (14).

An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68 percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using either Formula (9) or $p = 50$ in Formula (10), the standard error of an estimate of 50 percent of the group.

2. Add to and subtract from 50 percent, the standard error determined in step 1.

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group with more of the item is equal to the smaller

percentage found in step 2. This quantity will be the upper limit for the 68 percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group with more of the item is equal to the larger percentage found in step 2. This quantity will be the lower limit for the 68 percent confidence interval.

4. Divide the difference between the two quantities determined in step 3 by two to obtain the standard error of the median.

To perform step 3, it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Note, however, that Pareto interpolation can never be used if the interval contains zero or negative measures of the item of interest. Interpolation is used as follows.

The quantity of the item such that p percent have more of the item is:

$$X_{pN} = A_1 \times \exp\left[\left(\frac{\ln\left(\frac{qN}{N_1}\right)}{\ln\left(\frac{N_2}{N_1}\right)}\right)\ln\left(\frac{A_2}{A_1}\right)\right] \tag{13}$$

if Pareto Interpolation is indicated and:

$$X_{pN} = \left[A_1 + \left(\frac{qN - N_1}{N_2 - N_1}\right)(A_2 - A_1)\right], \tag{14}$$

if linear interpolation is indicated, where:

N	is the size of the group,
A_1 *and* A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls
N_1 *and* N_2	are the estimated number of group members owning more than A_1 and A_2, respectively
exp	refers to the exponential function and
ln	refers to the natural logarithm function
q	is the corresponding proportion for p i.e., $q = p/100$

Illustration 6.
To illustrate the calculations for the sampling error on a median, we return to Table A from illustration 2. The median annual income for this group using Formula (13) is $31,828. The size of the group is 23,527,000.

1. Using Formula (10), the standard error of 50 percent on a base of 23,527,000 is about 0.77 percentage points.

2. Following step 2, the two percentages of interest are 49.23 and 50.77.

3. By examining Table A, we see that the percentage 49.23 falls in the income interval from $30,000 to $39,999. (Since 54.7 percent receive more than $30,000 per annum, the dollar value corresponding to 49.23 must be between $30,000 and $40,000.) Thus, $A_1 = \$30{,}000, A_2 = \$40{,}000, N_1 = 12{,}881{,}000\ and\ N_2 = 8{,}285{,}000.$

In this case, we decided to use Pareto interpolation. Therefore, using Formula (13), the upper bound of a 68 percent confidence interval for the median is

$$\$30{,}000 \times \exp\left[\left(\frac{\ln\left(\frac{0.4923 \times 23{,}527{,}000}{12{,}881{,}000}\right)}{\ln\left(\frac{8{,}285{,}000}{12{,}881{,}000}\right)}\right) \times \ln\left(\frac{40{,}000}{30{,}000}\right)\right] = \$32{,}152.$$

Also, by examining Table A, we see that 50.77 falls in the same income interval. Thus, $A_1, A_2, N_1 and\ N_2$ are the same. We also use Pareto interpolation for this case. So the lower bound of a 68 percent confidence interval for the median is

$$\$30{,}000 \times \exp\left[\left(\frac{\ln\left(\frac{0.5077 \times 23{,}527{,}000}{12{,}881{,}000}\right)}{\ln\left(\frac{8{,}285{,}000}{12{,}881{,}000}\right)}\right) \times \ln\left(\frac{40{,}000}{30{,}000}\right)\right] = \$31{,}513.$$

Thus, the 68 percent confidence interval on the estimated median is from $31,513 to $32,152.

4. Then the approximate standard error of the median is

$$\frac{\$32{,}152 - \$31{,}513}{2} = \$319.5$$

3.11 Standard Errors of Ratios of Means and Medians

The standard error for a ratio of means or medians is approximated by:

$$s_{\frac{x}{y}} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]}, \quad (15)$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula (15) assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

3.12 Standard Errors Using Software Packages

Standard errors and their associated variance, calculated by statistical software packages such as SAS or Stata, do not accurately reflect the SIPP's complex sample design. Erroneous conclusions will result if these standard errors are used directly. We provide adjustment factors by characteristics that should be used to correctly compensate for likely under-estimates. The factors called design effects (DEFF), available in Table 6, must be applied to SAS or Stata generated variances. The square root of DEFF can be directly applied to similarly generated standard errors. These factors approximate design effects which adjust statistical measures for sample designs more complex than simple random sample.

Replicate weights for SIPP are also provided and can be used to estimate more accurate standard errors and variances. While replicate weighting methods require more computing resources, many statistical software packages, including SAS, have procedures that simplify the use of replicate weights for users. To calculate variances using replicate weights use the formula:

$$Var(\theta_0) = \frac{1}{G(0.5)^2} \times \sum_{i=1}^{G} (\theta_i - \theta_0)^2 \qquad (16)$$

where G is the number of replicates, θ_0 is the estimate using full sample weights, and θ_i is the estimate using the replicate weights. For the 2018 panel, G=240 for the number of replicate weights provided in the public use files. Replicate weights are created using Fay's method, with a Fay coefficient of 0.5 (Chakrabarty, 1993; Fay, 1984).

Instead of direct computation, various SAS procedures include options to use replicate weights when estimating standard errors or variances. To use Balanced Repeated Replication (BRR) method using replicate weights in SAS include the VARMETHOD=BRR(FAY=0.5) option in the PROC statement and specify the replicate weights with a REPWEIGHTS statement. Other computer packages have similar methods.
Formula (16) produces variance estimates close to zero for the median when multiple observations have value equal to the median. In this case, two methods can be used to estimate the variance of the median. The first technique incorporates replicate weights in Woodruff's method for estimating variability (Woodruff, 1952). Gossett et al. (2002) documents the procedure for combining Woodruff's method with Jackknife replication and provides sample codes adapted by Mack and Tekansik (2011) for Fay's BRR. The second method uses

VARMETHOD=TAYLOR option, a direct application of Woodruff's method, along with the cluster and strata statements instead of replicate weights to account for SIPP's complex design.

Illustration 7.

In SAS, the SURVEYMEANS procedure is used to estimate statistics such as means, totals, proportions, quantiles, and ratios for a survey sample. An example syntax for estimating the mean and median of the total household income, *THTOTINC*, for any month in 2017 using SIPP replicate weights is:

```
proc surveymeans data=pu2018w1[11] varmethod=brr(Fay=0.5) mean median[12];
  var THTOTINC;
  weight WPFINWGT;
  repweights REPWGT1-REPWGT240;
run;
```

Similarly, replicate weights can be used to estimate standard errors in the SURVEYFREQ (for frequency tables and cross-tabulations), SURVEYREG (for regression analysis), SURVEYLOGISTIC (for logistic regression analysis), and SURVEYPHREG (for proportional hazards regression analysis) SAS procedures by using the same VARMETHOD = BRR(FAY=0.5) option and REPWEIGHTS statement.

In Stata, the SVY command is used to fit a statistical model to a complex survey dataset. SVYSET is used to determine the survey design and provide information about the variance estimation. The following Stata syntax is equivalent to using SURVEYMEANS by SAS:

```
use pu2018w1.dta
svyset [pweight= WPFINWGT], brrweight(REPWGT1-REPWGT240) fay(.5) vce(brr) mse
svy: mean THTOTINC
```

[11] This snippet of code requires the analytic, final weight, and replicate weights variables – *THTOTINC, WPFINWGT, and REPWGT1-REPWGT240* respectively – are included in the dataset *pu2018w1.*

[12] The documentation for the Surveymeans procedure provides a list of "statistic keywords" that can be supplied to compute estimates of different statistics.

4. References

Chakrabarty, R. P. (1993). Variance estimation by users of SIPP micro-data files [Working Paper]. https://www.census.gov/content/dam/Census/library/working-papers/1993/adrm/rr93-6.pdf

Fay, R. E. (1984). Some properties of estimates of variance based on replication methods. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 495-500.

Gossett, J. M., Simpson, P., Parker, J.G., & Simon, W.L. (2002). How complex can complex survey analysis be with SAS? Proceedings of the 27th Annual SAS Users Group International Conference. https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p266-27.pdf

Kalton, G. (1998). SIPP quality profile [Working Paper]. U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/1998/demo/SEHSD-WP1998-11.pdf

Mack, S. & Tekansik, S. (2011). "Calculating Standard Errors and Confidence Intervals of Medians." U.S. Census Bureau Memorandum for Documentation dated May 6, 2011.

Sedgwick, P. (2014). Multiple hypothesis testing and Bonferroni's correction. BMJ: British Medical Journal, 349. https://www.jstor.org/stable/26517649

Stoline, M. (1981). The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. The American Statistician, 35(3), 134-141. https://www.jstor.org/stable/2683979

Tersine Jr., A.G. (2020a). "Survey of Income and Program Participation 2018: Weighting Specifications for Wave 1 (WGT-29)." U.S. Census Bureau Memorandum from Anthony G. Tersine, Jr. to Jeffrey Sission. Not yet finalized

Tersine Jr., A.G. (2020b). "Survey of Income and Program Participation Cross-sectional Weighting Specifications for Calendar Year 2019 and Subsequent Years (WGT-34)." U.S. Census Bureau Memorandum from Anthony G. Tersine, Jr. to Jeffery Sission. Not yet finalized

U.S. Census Bureau. (2019). Survey of Income and Program Participation: 2014 Panel Users' Guide. Washington, D.C. https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/2014-SIPP-Panel-Users-Guide.pdf

U.S. Census Bureau. (2020). Survey of Income and Program Participation: 2018 Panel Users' Guide. Washington, D.C. https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/2014-SIPP-Panel-Users-Guide.pdf

Wolter, K. M. (2007). Chapter 7: Generalized variance functions, Introduction to Variance Estimation, 2nd Ed. New York: Springer, pp. 272-297.

Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. Journal of the American Statistical Association, 47(260), 635-646.

5. Cross-Sectional Generalized Variance Parameters and Tables

Table 6. Generalized Variance Parameters for Wave 1

Domain	Parameters *a*	Parameters *b*	Design Effect[13]	*f*
Poverty and Program Participation, Persons 15+				
Total	-0.00002189	5,696	2.186	1.011
Male	-0.00004525	5,696		
Female	-0.00004240	5,696		
Income and Labor Force Participation, Persons 15+				
Total	-0.00002131	5,546	2.128	0.998
Male	-0.00004406	5,546		
Female	-0.00004128	5,546		
Other, Persons 0+				
Total (or White)	-0.00001733	5,568	2.137	1
Male	-0.00003546	5,568		
Female	-0.00003391	5,568		
Black, Persons 0+	-0.0001387	5,859	2.248	1.026
Male	-0.0002962	5,859		
Female	-0.0002608	5,859		
Hispanic, Persons 0+	-0.00009580	5,598	2.148	1.003
Male	-0.00019140	5,598		
Female	-0.00019190	5,598		
Households				
Total (or White)	-0.00003846	4,973	1.908	1
Black	-0.00028740	4,973		
Hispanic	-0.00027670	4,973		

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

Notes on Domain Usage for Table 6

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

[13] Design Effect=b/sample interval where sample interval=2,606

Table 7. Base Standard Errors of Estimated Numbers of Households or Families

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200,000	31,513	30,000,000	338,491
300,000	38,580	40,000,000	370,653
500,000	49,768	50,000,000	390,512
750,000	60,894	60,000,000	399,905
1,000,000	70,246	70,000,000	399,570
2,000,000	98,955	80,000,000	389,482
3,000,000	120,718	90,000,000	368,841
5,000,000	154,608	95,000,000	354,025
7,500,000	187,441	99,500,000	337,713
10,000,000	214,206	105,000,000	313,279
15,000,000	256,791	110,000,000	285,769
25,000,000	316,682	117,610,000	229,982

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

Notes:

1. These estimates are calculations using the Household Total (or White) a and b parameters from Table 6 and Formula (4).

Table 8. Base Standard Errors of Estimated Numbers of Persons

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200,000	33,360	110,000,000	634,655
300,000	40,851	120,000,000	646,999
500,000	52,723	130,000,000	656,478
750,000	64,547	140,000,000	663,213
1,000,000	74,503	150,000,000	667,289
2,000,000	105,198	160,000,000	668,754
3,000,000	128,639	170,000,000	667,625
5,000,000	165,550	180,000,000	663,889
7,500,000	201,953	190,000,000	657,501
10,000,000	232,265	200,000,000	648,383
15,000,000	282,171	210,000,000	636,417
25,000,000	358,286	220,000,000	621,440
30,000,000	389,157	230,000,000	603,227
40,000,000	441,579	240,000,000	581,474
50,000,000	484,845	250,000,000	555,765
60,000,000	521,241	260,000,000	525,521
70,000,000	552,126	270,000,000	489,901
80,000,000	578,384	275,000,000	469,701
90,000,000	600,622	280,000,000	447,625
100,000,000	619,274	299,340,000	337,461

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

Notes:

1. These estimates are calculations using the Other Persons 0+ a and b parameters from Table 6 and Formula (4).

2. Multiply the standard error from this table by the appropriate f factor from Table 6 to calculate the standard error for another domain and/or reference period. For example, to calculate standard error for Wave 1 cross-sectional estimates related to labor force characteristics, multiply the appropriate standard error (based on the size of the estimate) by $f = 0.998$

Table 9. Base Standard Errors for Percentages of Households or Families

Base of Estimated Percentages	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200,000	1.57	2.21	3.44	4.73	6.83	7.88
300,000	1.28	1.80	2.81	3.86	5.58	6.44
500,000	0.99	1.40	2.17	2.99	4.32	4.99
750,000	0.81	1.14	1.77	2.44	3.53	4.07
1,000,000	0.70	0.99	1.54	2.12	3.05	3.53
2,000,000	0.50	0.70	1.09	1.50	2.16	2.49
3,000,000	0.41	0.57	0.89	1.22	1.76	2.04
5,000,000	0.31	0.44	0.69	0.95	1.37	1.58
7,500,000	0.26	0.36	0.56	0.77	1.12	1.29
10,000,000	0.22	0.31	0.49	0.67	0.97	1.12
15,000,000	0.18	0.25	0.40	0.55	0.79	0.91
25,000,000	0.14	0.20	0.31	0.42	0.61	0.71
30,000,000	0.13	0.18	0.28	0.39	0.56	0.64
40,000,000	0.11	0.16	0.24	0.33	0.48	0.56
50,000,000	0.10	0.14	0.22	0.30	0.43	0.50
60,000,000	0.09	0.13	0.20	0.27	0.39	0.46
70,000,000	0.08	0.12	0.18	0.25	0.36	0.42
80,000,000	0.08	0.11	0.17	0.24	0.34	0.39
90,000,000	0.07	0.10	0.16	0.22	0.32	0.37
105,000,000	0.07	0.10	0.15	0.21	0.30	0.34
110,000,000	0.07	0.09	0.15	0.20	0.29	0.34
117,610,000	0.06	0.09	0.14	0.20	0.28	0.33

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

Notes:

1. These estimates are calculations using the Households Total (or White) b parameter from Table 6 and Formula (10).

Table 10. Base Standard Errors for Percentages of Persons

Base of Estimated Percentages	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200,000	1.66	2.34	3.64	5.01	7.22	8.34
300,000	1.36	1.91	2.97	4.09	5.90	6.81
500,000	1.05	1.48	2.30	3.17	4.57	5.28
750,000	0.86	1.21	1.88	2.58	3.73	4.31
1,000,000	0.74	1.04	1.63	2.24	3.23	3.73
2,000,000	0.52	0.74	1.15	1.58	2.28	2.64
3,000,000	0.43	0.60	0.94	1.29	1.87	2.15
5,000,000	0.33	0.47	0.73	1.00	1.44	1.67
7,500,000	0.27	0.38	0.59	0.82	1.18	1.36
10,000,000	0.23	0.33	0.51	0.71	1.02	1.18
15,000,000	0.19	0.27	0.42	0.58	0.83	0.96
25,000,000	0.15	0.21	0.33	0.45	0.65	0.75
30,000,000	0.14	0.19	0.30	0.41	0.59	0.68
40,000,000	0.12	0.17	0.26	0.35	0.51	0.59
50,000,000	0.10	0.15	0.23	0.32	0.46	0.53
60,000,000	0.10	0.13	0.21	0.29	0.42	0.48
70,000,000	0.09	0.12	0.19	0.27	0.39	0.45
100,000,000	0.07	0.10	0.16	0.22	0.32	0.37
110,000,000	0.07	0.10	0.16	0.21	0.31	0.36
120,000,000	0.07	0.10	0.15	0.20	0.29	0.34
130,000,000	0.07	0.09	0.14	0.20	0.28	0.33
140,000,000	0.06	0.09	0.14	0.19	0.27	0.32
150,000,000	0.06	0.09	0.13	0.18	0.26	0.30
160,000,000	0.06	0.08	0.13	0.18	0.26	0.29
170,000,000	0.06	0.08	0.12	0.17	0.25	0.29
180,000,000	0.06	0.08	0.12	0.17	0.24	0.28
190,000,000	0.05	0.08	0.12	0.16	0.23	0.27
200,000,000	0.05	0.07	0.11	0.16	0.23	0.26
210,000,000	0.05	0.07	0.11	0.15	0.22	0.26
220,000,000	0.05	0.07	0.11	0.15	0.22	0.25
230,000,000	0.05	0.07	0.11	0.15	0.21	0.25
240,000,000	0.05	0.07	0.10	0.14	0.21	0.24
250,000,000	0.05	0.07	0.10	0.14	0.20	0.24
280,000,000	0.04	0.06	0.10	0.13	0.19	0.22
299,340,000	0.04	0.06	0.09	0.13	0.19	0.22

Source: U.S. Census Bureau, 2018 Survey of Income and Program Participation

Notes:

1. These estimates are calculations using the Other Persons 0+ a and b parameter from Table 6 and Formula (10).

2. Multiply the standard error from this table by the appropriate f factor from Tables 6 to calculate the standard for another domain and/or reference period.